STUDIES ON MONOTREME PROTEINS

I. AMINO ACID SEQUENCE OF THE β-CHAIN OF HAEMOGLOBIN FROM THE ECHIDNA, *TACHYGLOSSUS ACULEATUS ACULEATUS*

By R. G. WHITTAKER,* W. K. FISHER,* and E. O. P. THOMPSON*

Abstract

Blood from the echidna (or spiny anteater), *T. aculeatus aculeatus*, contained two haemoglobins which were separated by chromatography on DEAE-Sephadex. The major component, Hb-I, was converted to globin and fractionated into the β- and α-chains by chromatography on carboxymethyl cellulose in 8M urea–thiol buffers.

The complete amino acid sequence of the 146 residues in the β-chain has been determined. Cleavage with cyanogen bromide gave two polypeptides. Peptides derived from these or from the β-globin by tryptic digestion were isolated by gel filtration, paper ionophoresis, and chromatography. The amino acid sequences were determined by the "dansyl"–Edman procedure or by further digestion with other enzymes.

The amino acid sequence shows 31 changes compared with human β-chain, and consistently less differences from the sequences of β-globins of eutherian mammals than was the case with β-globins of marsupials. A small variation in the highly conserved residues of the G helix has not so far been reported for other haemoglobins. The amino acid residues in this β-chain which are in contact with the haem group or the α-chains are almost identical to the residues known to be involved in these interactions in horse haemoglobin.

I. INTRODUCTION

The echidna (or spiny anteater), *Tachyglossus aculeatus aculeatus* Shaw & Nodder, 1792, is one of two surviving species of the subclass Prototheria (class Mammalia). These monotremes have characteristic features in their anatomy, reproduction, and metabolism that relate them to both reptiles and higher mammals but the fossil record is too inadequate to be specific about their evolutionary development and radiation (Griffiths 1968).

A biochemical approach to this problem involves a study of the amino acid sequences of proteins common to many species (Zuckerkandl and Pauling 1962; Dayhoff 1969). In previous papers (Thompson, Hosken, and Air 1969; Air and Thompson 1969, 1971; Beard and Thompson 1971; Thompson and Air 1971) the amino acid sequences of the α-, β-, and myoglobin chains of marsupials have been presented and the data has been used (Air *et al.* 1971) to calculate a date of

* School of Biochemistry, University of New South Wales, P.O. Box 1, Kensington, N.S.W. 2033.

marsupial–eutherian divergence. The haemoglobins in the blood of the echidna have been examined by similar methods and the amino acid sequence of the β-globin present in the major haemoglobin is presented.

II. Materials and Methods

The methods of digestion with trypsin, thermolysin, pepsin, and papain, high-voltage ionophoresis, peptide mapping, amino acid analysis, and sequence determination by the dansyl–Edman procedure were the same as previously described (Air and Thompson 1969, 1971; Beard and Thompson 1971).

(a) *Fractionation of Haemolysate*

Samples of blood were freed from plasma and haemolysed as described by Thompson, Hosken, and Air (1969), with the addition of sufficient 3M Tris.HCl at pH 8·5 to make the final solution 0·01M; otherwise, the haemoglobins tended to precipitate. The cell debris was washed once with one volume of water and the combined solution converted to the carbon monoxide form.

Starch-gel electrophoresis of haemolysates of echidna blood has been studied by Cooper and Griffiths (personal communication, 1971). Polymorphic forms have been identified as A, B, and C major and A and B minor, based on their mobilities. Examination by starch-gel electrophoresis in gels containing 2M urea, 0·01M KCN, and 0·01M EDTA of the haemolysate used in this work showed the presence of two electrophoretic components (B major and A minor). These were separated on a column of DEAE-Sephadex (2·8 by 16 cm), equilibrated with 0·05M Tris.HCl containing 100 mg KCN per litre at pH 8·0 (Dozy, Kleihauer, and Huisman 1968), and eluted with similar strength buffer of gradually decreasing pH to a limit of pH 7·5 (linear gradient device with 300 ml each chamber). One component was eluted rapidly and the more tightly bound component could then be eluted with limit buffer containing added sodium chloride to hasten its elution.

The more tightly bound component was the major component called Hb-I (B major electrophoretically), which migrated least rapidly during starch-gel electrophoresis at pH 8·6 and was in approximately three times the concentration of Hb-II (A minor electrophoretically) in the blood.

(b) *Fractionation of Globin-I*

The globin was prepared from the Hb-I and fractionated as described by Thompson, Hosken, and Air (1969). The samples were chromatographed on carboxymethyl cellulose columns according to the method of Clegg, Naughton, and Weatherall (1965) using a linear gradient from 250 ml each of 0·01M Na^+ (phosphate) and 0·05M Na^+ (phosphate) in 8M urea containing disodium EDTA (0·001M) and mercaptoethanol (0·05M). The column was equilibrated at the lower salt concentration. The globin chains were carboxymethylated before recovery by dialysis and freeze-drying.

(c) *Cyanogen Bromide Cleavage*

The *S*-carboxymethylated β-globin (50 mg) was dissolved in 70% formic acid (1 ml) and reacted with cyanogen bromide (50 mg) for 24 hr at 4°C or room temperature. The reaction mixture was freeze-dried, redissolved in 0·35 ml of formic acid, diluted with 0·35 ml water, and loaded on a column (1 cm diam. by 140 cm) of Sephadex G75 equilibrated and freshly washed with 10% formic acid. Fractions of 1·5 ml were collected and their absorption at 280 nm measured. The pooled fractions were recovered after dilution with water and freeze-drying.

(d) *Fractionation of Tryptic Peptides*

(i) *Gel Filtration*

The SCM-globin was not completely solubilized by 4 hr tryptic digestion at 37°C and some insoluble peptides plus partially degraded β-chain were removed by centrifugation. The soluble fraction was loaded on a Sephadex G50 column (2·8 cm diam. by 120 cm), equilibrated, and

eluted with 1% ammonium bicarbonate, pH 9. The extinction of the eluate at 280 nm was measured and four fractions were bulked and freeze-dried. Each fraction was dissolved in pH 6·4 pyridine acetate buffer (4 ml) and any insoluble peptides centrifuged off and washed with fresh buffer. Each soluble fraction was examined by peptide mapping and a sample of each insoluble fraction was hydrolysed for amino acid analysis.

(ii) *Peptides Insoluble at pH 8·7*

During tryptic digestion of β-globin at pH 8·7 some peptide material may have precipitated but it was contaminated with partially degraded β-chain. The main insoluble peptide was from the CNBr2 fraction of β-chain and this was isolated from CNBr2 after digestion with TPCK-trypsin overnight at 37°C. The digest was freeze-dried and peptides soluble in pH 6·4 pyridine acetate buffer extracted, followed by centrifugation and washing with 0·2M *N*-ethyl morpholine acetate buffer, pH 8·5 then 0·2M pyridine acetate buffer, pH 3·1.

A further sample was isolated from 25 mg CNBr2 after citraconylation (Dixon and Perham 1968). The protein was first dissolved in water (1 ml) and diluted to 5 ml with 0·2M sodium borate, pH 9. Citraconic anhydride (0·1 ml) was added and the pH of the magnetically stirred solution kept constant by the addition of 5M NaOH as required. The citraconyl protein was dialysed against water, centrifuged free of any insoluble material, and freeze-dried before digestion with trypsin for 3 hr at 37°C. Only a very small amount of material precipitated during digestion and the digest was freeze-dried. The soluble citraconyl peptides were dissolved in pH 6·4 pyridine acetate buffer and the residue washed three times with the same buffer. The insoluble material was then washed three times with pH 3·1 pyridine acetate buffer.

The soluble citraconylated peptides were fractionated on a Sephadex G50 column, 130 cm by 1 cm diameter, and eluted with 1% ammonium bicarbonate.

III. Results

The nomenclature of Gerald and Ingram (1961) has been used to describe the β-chain tryptic peptides. The numbering of peptides from the *N*-terminal end follows that of human β-chain as far as possible. The symbols A and B have been

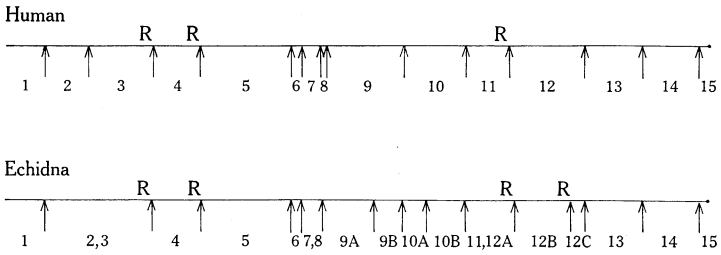


Fig. 1.—Nomenclature for β-chain tryptic peptides of echidna haemoglobin-I(B) compared with β-chain tryptic peptides of human haemoglobin. The diagrams are scaled to show the relative numbers of the residues in each peptide. ↑ Lysine residue. R↑ Arginine residue.

added where an extra arginine or lysine residue has occurred in the echidna chain and when a lysine has not occurred in the echidna chain the peptide has been given all the numbers of the corresponding human peptides (see Fig. 1).

As in previous papers the CNBr fragments are numbered from the *N*-terminus CNBr1 and CNBr2.

(*a*) *Fractionation of Soluble Tryptic Peptides*

The peptide map of the tryptic peptides soluble at pH 6·4 when separated by paper ionophoresis–chromatography is shown in Figure 2. Separation of tryptic

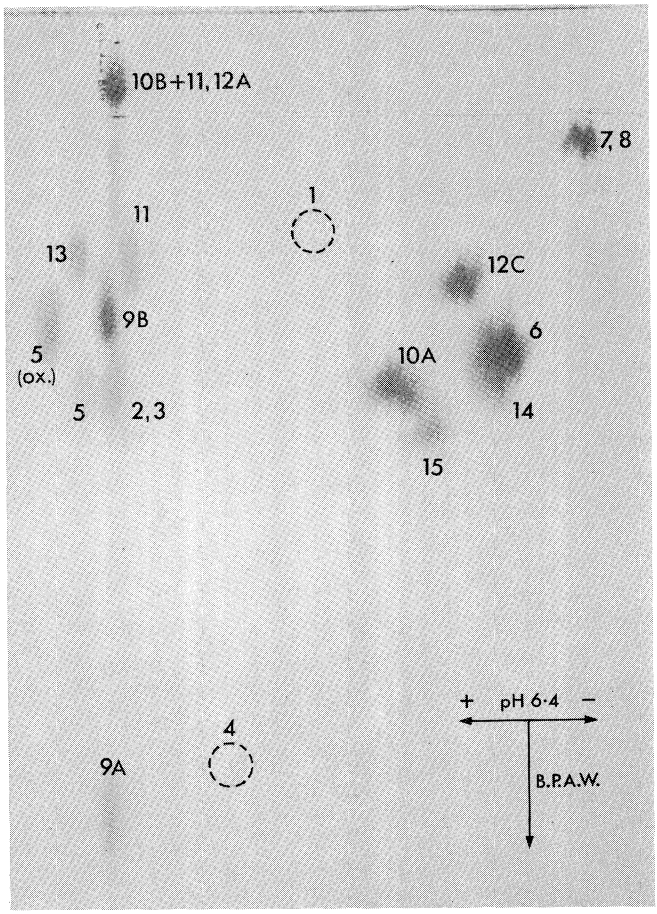


Fig. 2.—Peptide map of β-chain of echidna Hb-I(B) globin digested with TPCK-trypsin. Ionophoresis at pH 6·4 was followed by chromatography with butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v) as indicated. The peptides are given the identification number corresponding to their position in the chain as shown in Figure 1.

peptides by gel filtration on Sephadex G50 is shown in Figure 3. The peptides present in each fraction and the peptides insoluble at pH 6·4 are as follows:

Fraction 1	Fraction 2	Fraction 3	Fraction 4
	Peptides soluble at pH 6·4		
βTp2+3	βTp1	βTp1*	βTp6
βTp5	βTp5	βTp4	βTp7,8*
βTp10B+11,12A	βTp9A	βTp7,8	βTp10A*
	βTp9B	βTp9B*	βTp12C*
	βTp11,12A	βTp9A*	βTp15
	βTp13	βTp10A	
		βTp11,12A*	
		βTp12C	
		βTp13*	
		βTp14	
	Peptides insoluble at pH 6·4		
βTp2,3*	βTp2,3		

* Indicates major proportion of peptide is in another fraction.

Peptides in the soluble portion of each fraction were purified using paper techniques. The soluble peptides included all tryptic peptides from the β-chain except βTp2,3 and βTp12B. Peptide βTp2,3 was the main insoluble peptide in fraction 2 from the Sephadex column while peptide βTp12B was precipitated during the tryptic digestion.

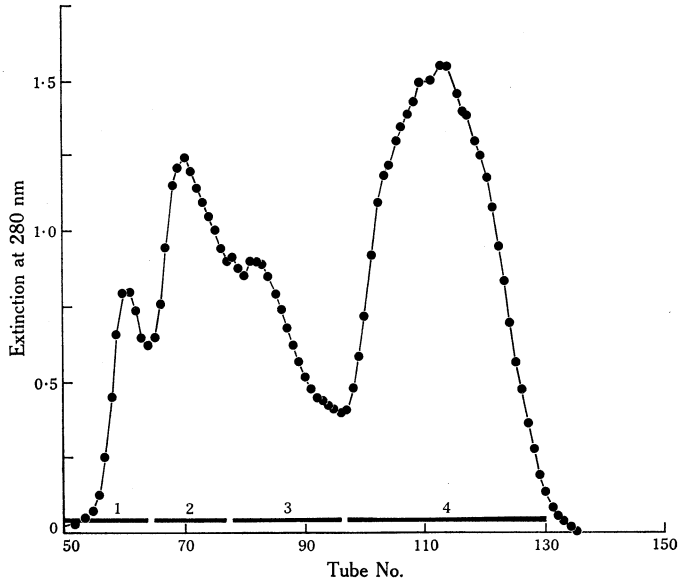


Fig. 3.—Gel filtration of the peptides soluble at pH 8·7 after tryptic digestion of the β-chain of echidna globin (500 mg) on a column (110 by 2·8 cm diam.) of Sephadex G50 in 1% ammonium bicarbonate, pH 9·1. Flow rate 36 ml/hr; fraction size 4·8 ml. Tubes containing fractions combined for peptide recovery are shown as bars.

(*b*) *Cyanogen Bromide Cleavage of Echidna β-I Globin*

The amino acid analysis of echidna β-I globin (Table 1) showed one methionine residue, hence cyanogen bromide cleavage would be expected to yield only two fragments. The results of gel filtration of the reaction mixture after 24 hr treatment at room temperature in 70% formic acid is shown in Figure 4. There were three main fractions including some unchanged β-globin, possibly containing methionine sulphoxide which would prevent cleavage. The results were similar at 4°C and each fraction was re-run on the same column to give peak material that on amino acid analysis gave the values shown in Table 1.

The amino acid analyses and the peptide maps of tryptic digests showed that the methionine was residue 55 as in other β-globins.

(*c*) *Isolation of Insoluble-core Peptide βTp12B*

The isolation of βTp12B as a pure peptide was difficult, in line with the problems reported by other workers who encountered a similar peptide in their work. Babin *et al.* (1966) working with the insoluble material from a tryptic digest of foetal

calf β-chain and Schroeder *et al.* (1967) with the material from bovine β-chains obtained their best purification by gel filtration in 50% acetic acid. When the insoluble material from a tryptic digest of echidna β-chain, even after redigestion for 16 hr, was subjected to gel filtration on Sephadex G50 in 10% formic acid a series of peak

TABLE 1

AMINO ACID COMPOSITIONS OF ECHIDNA β-GLOBIN AND ITS CNBr FRAGMENTS

The proteins were purified as described in the text then hydrolysed for 24 hr under vacuum with 6N HCl containing 1 mg phenol per millilitre. Values are given as mole per mole of each protein or fragment, uncorrected for losses on hydrolysis or incomplete hydrolysis. Also shown are the amino acid compositions obtained from the amino acid sequence

Amino acid	Echidna β-IB globin		CNBr1 fragment		CNBr2 fragment	
	Hydrolysate*	Sequence	Hydrolysate†	Sequence	Hydrolysate†	Sequence
Lysine	11·1	11	1·5	1	9·5	10
Histidine	8·9	9	2·2	2	6·7	7
Arginine	4·0	4	2·0	2	2·1	2
SCM-cysteine	1·0	1	0	0	0·8	1
Aspartic acid	15·2	15	4·9	5	10·4	10
Threonine‡	4·1	6	2·5	3	2·9	3
Serine‡	3·6	10	3·7	5	4·5	5
Glutamic acid	12·4	11	5·1	5	6·6	6
Proline	3·6	3	1·0	1	2·1	2
Glycine	12·5	12	5·7	6	6·4	6
Alanine	16·1	15	4·3	4	11·0	11
Valine	16·4	16	6·4	7	8·6¶	9
Methionine	0·9	1	0	0	0	0
Leucine	19·6	19	6·8	7	12·2	12
Tyrosine	2·1	2	0·9	1	1·0	1
Phenylalanine	8·3	8	2·8	3	4·9	5
Tryptophan§	2·9	3	1·4	2	0·7	1
Homoserine ‖			(+)	1		
Total		146		55		91

* Hydrolysed at 115–120°C. At this temperature destruction of serine and threonine was much higher while liberation of valine was more complete.

† Hydrolysed at 110°C.

‡ Uncorrected for decomposition.

§ Alkaline hydrolysis (Neumann, Moore, and Stein 1962) for β-globin and with thioglycollic acid added to the HCl (Matsubara and Sasaki 1969) for the CNBr fragments.

‖ The hydrolysates were treated with 0·1M K_2CO_3 (Thompson and O'Donnell 1967) to convert homoserine lactone to homoserine before analysis. For these samples the ion-exchange columns were eluted initially with pH 3·07 buffer (cf. Ambler 1965).

¶ This rose to 9·3 residues after 96 hr hydrolysis.

fractions were obtained but all analysed as mixtures with βTp12B as a major component contaminated with βTp2,3, βTp4, and βTp13. The two former contaminants would not be present in the larger CNBr2 fragment comprising residues 56–146.

Tryptic digests of CNBr2 gave a precipitate which after washing with pyridine acetate buffer at pH 6·4 and *N*-ethylmorpholine acetate buffer at pH 8·5 still

analysed as a mixture containing contaminants. Following the recommended procedure of Boyer *et al.* (1967) this peptide material was washed several times with pyridine acetate buffer, pH 3·1, to give material showing a lower level of contamination. A better product was obtained by taking advantage of the solubilizing effect of citraconyl groups when substituted on all free amino groups (Dixon and Perham 1968) and the presence of susceptible arginyl bonds linking peptide βTp12B to the

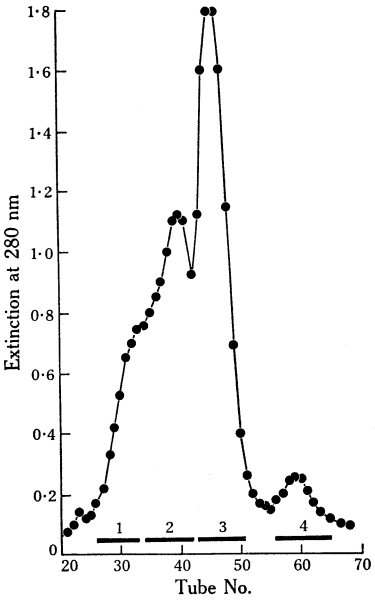


Fig. 4.—Gel filtration of cyanogen bromide fragments from echidna β-globin on a column of Sephadex G50 (140 by 1 cm diam.) eluted with 10% formic acid. Fraction size 1·5 ml; the fractions pooled for analyses are indicated by bars. Fraction 4 contained no peptide material.

rest of the chain. The insoluble material from a 3-hr tryptic digest analysed as the purest peptide preparation with a composition similar to that found by Babin *et al.* (1966) and Schroeder *et al.* (1967) for bovine foetal β-chain and bovine β-chain respectively, and Boyer *et al.* (1967) for sheep β-chains, although amino acid analysis and sequence data indicated one less residue of leucine.

(*d*) *Amino Acid Composition of Tryptic Peptides*

The amino acid compositions of purified peptides are given in Table 2.

The total compositions are in good agreement with the analysis of the β-chain (Table 1) except the values of glutamic acid and alanine which are high in the globin analysis. The analysis for CNBr1 is in excellent agreement with the sequence data, while that for CNBr2 again shows a somewhat higher value for glutamic acid and a lower value for valine. The presence of a Val-Val-Val sequence in βTp12B probably accounts for the lower valine values since at least 96 hr hydrolysis is required for complete hydrolysis of Val-Val peptide bonds. The glutamic acid values in our laboratory tend to be high, probably due to loss of glutamic acid from, or a low content in, the standard mixture.

The soluble citraconylated peptides Cit-CNBr2-Tp1, corresponding to the *N*-terminal portion of CNBr2 (residues β56–105) and Cit-CNBr2-Tp3, corresponding

TABLE 2

AMINO ACID COMPOSITION OF ECHIDNA β-IB GLOBIN TRYPTIC PEPTIDES

Soluble peptides were purified by gel filtration, paper ionophoresis at pH 6·4, and paper chromatography. Hydrolysates were made at 110°C for 24 hr. Values are not corrected for losses during hydrolysis or due to incomplete hydrolysis and are given as moles per mole of peptide, with preferred values in parentheses. Detection by ninhydrin results in low recoveries of the *N*-terminal amino acid

Amino acid	βTp1	βTp2,3*	βTp4	βTp5	βTp6	βTp7,8	βTp9A	βTp9B	βTp10A	βTp10B† +11,12A	βTp12B‡	βTp12C	βTp13	βTp14	βTp15	Total in protein
Lysine	1·0(1)	0·2		1·0(1)	1·0(1)	1·0(1)	1·0(1)	1·0(1)	1·0(1)	1·0(1)		1·0(1)	0·9(1)	1·0(1)		11
Histidine	0·9(1)	0·9(1)				0·8(1)				1·8(2)		0·6(1)		2·1(2)	1·0(1)	9
Arginine		0·9(1)	0·9(1)							0·9(1)	1·0(1)					4
SCM-cysteine										0·5(1)						1
Aspartic acid		3·2(3)		2·6(3)			0·7(1)	2·8(3)		4·0(4)	1·1(1)		0·3			15
Threonine		1·6(2)	1·2(1)				0·7(1)		0·9(1)				0·9(1)			6
Serine	1·7(2)	0·4		2·7(3)			0·8(1)			1·3(1)		0·9(1)		1·7(2)		10
Glutamic acid	1·0(1)	2·0(2)	1·3(1)	1·2(1)						2·2(2)			4·0(4)			11
Proline			1·2(1)							0·9(1)			0·9(1)			3
Glycine	1·1(1)	3·8(4)	0·3	2·0(2)		1·1(1)	1·1(1)		0·6(1)	0·5	0·9(1)			1·0(1)		12
Alanine		2·1(2)		2·7(3)		1·8(2)	1·1(1)		1·1(1)	0·4	0·9(1)		3·0(3)	2·0(2)		15
Valine	0·8(1)	3·1(3)	2·0(2)	1·0(1)	1·0(1)		1·0(1)			1·1(1)	3·9(4)			2·0(2)		16
Methionine				0·6(1)												1
Leucine	1·0(1)	3·2(3)	1·6(2)	1·0(1)			2·1(2)	2·0(2)		2·6(3)	3·1(3)			1·7(2)		19
Tyrosine			1·2(1)												0·4(1)	2
Phenylalanine		0·3		2·0(3)			1·0(1)		1·0(1)	1·0(1)		1·0(1)	1·1(1)			8
Tryptophan§		0·5(1)	+(1)										+(1)			3
Total	8	22	10	19	2	5	10	6	5	18	11	4	12	12	2	146

* This peptide is insoluble at pH 6·4 but soluble at pH 8·7. Isolated from gel-filtration fraction 2 (Fig. 2).

† This is the major peptide containing Tp10B and Tp11,12A due to incomplete cleavage.

‡ This peptide precipitates during tryptic digestion at pH 8·7. Purified from citraconyl-CNBr2 tryptic digest and hydrolysed 96 hr. Corrected for 8% contamination with undigested citraconyl-CNBr2.

§ Detected by Ehrlich reagent.

to the *C*-terminal sequence β117–146, were partially separated on Sephadex G50 columns. Amino acid analyses of the fractions supported the allocation of tryptic peptides made to these areas of the β-chain.

(*e*) *Amino Acid Sequences of Tryptic Peptides*

The results of degradations by the dansyl–Edman procedure have been sufficiently clear cut to leave no doubt as to the correct sequence. Amide group allocations have been made on the basis of peptide mobilities, often after further enzyme digestion of the tryptic peptide. As in previous papers (Air and Thompson 1969, 1971; Beard and Thompson 1971; Thompson and Air 1971), residues identified by the dansyl–Edman procedure have been printed in *italic* fount.

Each peptide is discussed in turn starting from the *N*-terminal end. The β-chain of echidna is so strongly homologous to human and kangaroo β-globin that the peptides have been aligned by homology and the evidence obtained to support this alignment will not be presented.

βTp1

The *N*-terminal sequence of the β-globin chain was shown to be *Val-His-Leu-Ser-* by the Edman method (Blombäch *et al.* 1966), and the complete sequence of the *N*-terminal tryptic peptide by dansyl–Edman degradation is

Val-His-Leu-Ser-Gly-Ser-Glu-Lys

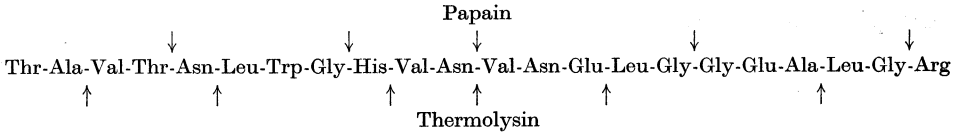
The presence of a glutamic acid residue was confirmed by the mobility during ionophoresis at pH 6·4 of the peptide and the neutral thermolysin fragment Leu-Ser-Gly-Ser-Glu-Lys.

βTp2,3

This large peptide was soluble at pH 8·7 but insoluble at pH 6·4. The sequence of this peptide is

*Thr-Ala-Val-Thr-Asn-Leu-*Trp-*Gly-*His-*Val-Asn-Val-Asn-Glu-Leu-Gly-Gly-Glu-Ala-Leu-Gly-*Arg

The dansyl–Edman procedure gave the sequence *Thr-Ala-Val-Thr-Asx*. Thermolysin and papain digestion gave a series of overlapping peptides (cf. Thompson, Sleigh, and Smith 1971; Beard and Thompson 1971) that established the sequence. The peptides were isolated by peptide mapping involving ionophoresis at pH 6·4 followed by chromatography in the second dimension in butanol–pyridine–acetic acid–water (15 : 10 : 3 : 12 v/v). The peptides identified in these experiments are listed in Table 3, the compositions of the papain peptides were identified by ionophoresis of 6N HCl hydrolysates at pH 1·8. The points of cleavage by the enzymes are shown below:



βTp4

The peptide is basic at pH 6·4, therefore glutamine is present. The sequence was determined by the dansyl–Edman procedure as

*Leu-Leu–Val-Val-Tyr-Pro-*Trp-*Thr-Gln-Arg*

The peptide gave a positive test for tryptophan by the Ehrlich reaction but no residue was detected by the dansyl method at the seventh position.

TABLE 3

SEQUENCES OF THERMOLYSIN AND PAPAIN PEPTIDES FROM βTp2,3

Residues identified by the dansyl–Edman procedure are shown in *italics*

Thermolysin peptides	Papain peptides
Thr-Ala	Thr-Ala-Val-Thr
Val-Thr-Asn*	Asn-Leu-Trp-Gly
(neutral at pH 6·4, therefore asparagine)	(neutral at pH 6·4, therefore asparagine)
Leu-Trp-*Gly*-His	His-Val-Asn
(positive test for tryptophan)	(basic at pH 6·4, therefore asparagine
*Val-Asn**	Val-Asn-Glu-Leu-Gly
(neutral at pH 6·4, therefore asparagine)	(acidic at pH 6·4 with a net charge of −1)
Val-Asn-Glu†	Gly-Glu-Ala-Leu-Gly
Leu-Gly-Gly-Glu-Ala	(acidic at pH 6·4, therefore glutamic acid)
(acidic at pH 6·4, therefore glutamic acid)	Arginine
Leu-Gly-Arg	

* One spot contained both sequences shown. The second step of the dansyl–Edman procedure gave two dansyl-amino acids but only DANS-Asp in step 3. The two peptides were separated by further ionophoresis at pH 1·8 and identified after hydrolysis with 6N HCl.

† This peptide was acidic at pH 6·4 with a net charge of −1. After the second Edman cleavage the thiazolinone was converted to the PTH-amino acid in 1N HCl at 80°C for 10 min. An ethyl acetate extract contained PTH-asparagine while the aqueous phase showed glutamic acid rather than glutamine by paper ionophoresis at pH 3·5.

βTp5

This peptide, which was acidic at pH 6·4, was located in more than one position of the peptide map due to partial oxidation of the methionine residue which analysis showed was present.

The methionine residue was known to occur at position β55 from the analyses of CNBr1 and CNBr2. The *C*-terminal sequence of the peptide was also known from that of a basic tryptic peptide isolated from CNBr2 which had been sequenced as *Gly-Asn-Ala-Lys*. The dansyl–Edman degradation gave an *N*-terminal sequence

Phe-Phe-Glx-Ser-Phe-Gly-Asx-Leu-Ser-Ser-Ala-Asx-Ala

Digestion with thermolysin gave peptides consistent with this sequence and the specificity of the enzyme. Those identified by amino acid analysis of hydrolysates using paper ionophoresis at pH 1·8 included

Phe-Phe-Glu-Ser	Acidic at pH 6·4, therefore glutamic acid
Phe-Gly-Asp	Acidic at pH 6·4, therefore aspartic acid
Leu-Ser-Ser	
Ala-Asp-Ala	Acidic at pH 6·4, therefore aspartic acid
Ala-Asp-Ala-Val	Linking the *N*-terminal sequence above to next peptide
Val- Met-Gly-Asn	Neutral at pH 6·4, therefore asparagine
Ala-Lys	
Val-Met-Gly-Asn-Ala-Lys	

These peptides gave a complete sequence for βTp5 as shown below together with the points of hydrolysis by thermolysin:

Phe-Phe-Glu-Ser-Phe-Gly-Asp-Leu-Ser-Ser-Ala-Asp-Ala-Val-Met-*Gly-Asn-Ala-Lys*

↑ ↑ ↑ ↑ ↑ ↑

βTp6 ¦ ¦ ¦

The sequence of this peptide was *Val*-Lys

βTp7,8

This peptide gave a clear degradation by the dansyl–Edman procedure, the histidine being identified as PTH-His after hydrolysis of the thiazolinone:

Ala-His-Gly-Ala-Lys

βTp9A

This was a neutral peptide at pH 6·4, therefore aspartic acid rather than asparagine is present. The dansyl–Edman procedure gave the sequence

Val-Leu-Thr-Ser-Phe-Gly-Asp-Ala-Leu-Lys

βTp9B

A neutral peptide so that two asparaginyl and one aspartic acid residue must be present. The PTH-derivatives were isolated and identified at steps 1, 3, and 4 of the dansyl–Edman procedure to give the sequence

Asn-Leu-Asp-Asn-Leu-Lys

βTp10A

This peptide gave a yellow colour with ninhydrin and was readily sequenced as

Gly-Thr-Phe-Ala-Lys

βTp10B+11,12A

This is the major peptide containing βTp10B and βTp11,12A due to incomplete cleavage of an -Asp-Lys-Leu- bond. Some of the separate peptide βTp11,12A was also isolated. The analysis of a sample of βTp11,12A showed

$His_{1.0}Arg_{0.9}Asp_{3.3}Glu_{1.4}Pro_{1.1}Val_{1.1}Leu_{0.8}Phe_{1.0}$

The sequences determined on both βTp10B+11,12A and βTp11,12A by the dansyl–Edman procedure were in agreement with the sequence

Leu-Ser-Glx-Leu-His-*SCMCys-Asx*-Lys-*Leu*-His-*Val-Asx-Pro-Glx-Asx-Phe-Asx-Arg*

Confirmatory evidence for this sequence and for the identification of Asp or Asx and Glu or Glx residues was obtained from thermolysin digests of the βTp10B+11,12A. The peptides isolated were Leu-Ser-Glu, acidic at pH 6·4; Leu-His-SCMCys-Asp-Lys, neutral at pH 6·4; Leu-His; Val-Asp-Pro-Glu-Asn, strongly acidic at pH 6·4 with higher mobility than Leu-Ser-Glu and *C*-terminal asparagine by carboxypeptidase digestion, as well as by identification of asparagine by paper ionophoresis at pH 3·5 after the removal of four amino acids by the Edman degradation; *Phe-Asn-Arg*, basic at pH 6·4.

The complete sequence showing the bonds hydrolysed by thermolysin is

Leu-Ser-Glu-Leu-His-*SCMCys-Asp*-Lys-*Leu*-His-*Val-Asp-Pro-Glu-Asn-Phe-Asn-Arg*

↑ ↑ ↑ ↑

The insertion of an extra residue in this peptide compared with the similar peptide from other β-globins is confirmed by the analysis of βTp10B+11,12A and the formation of *Phe-Asn-Arg* on thermolysin digestion compared with Phe-Lys from the corresponding Tp10B+11 peptide from the potoroo (Thompson and Air 1971). The peptide βTp10B+11,12A from Hb-I(B) was obtained from three different animals and all gave the same amino acid composition.

βTp12B

The difficulties in isolating pure samples of this peptide have already been discussed. The samples obtained after extended times of trypsin digestion, 8 hr at 37°C or for longer times, seemed to have a lower leucine content than the corresponding peptide from bovine foetal β-globin (Babin *et al.* 1966), bovine β-chain (Schroeder *et al.* 1967) or sheep β-chain (Boyer *et al.* 1967). This peptide is hydrophobic and partially insoluble in 60% pyridine. The first dansyl amino acid was only leucine but in subsequent steps two amino acids appeared as dansyl derivatives, i.e. step 2 gave mainly dansylglycine with some dansylleucine which probably is due to incomplete coupling of the insoluble material in the first step. Step 3 gave dansylaspartic acid plus dansylglycine, etc. From a 3- or 8-hr tryptic digest of citraconyl-CNBr2 the peptide analysed for three leucine residues. The dansyl–Edman procedure gave the sequence

Leu-Gly-Asx-Val-Leu-Val-Val-Val-Leu-Ala-Arg

Dansylvalylvaline was present at steps 6 and 7 as a major product.

When the peptide was digested with pepsin and fractionated by the usual peptide mapping procedure only one basic peptide, Ala-Arg, was obtained. The neutral peptides contained a peptide (Leu,Gly,Asp,Val) indicating that the aspartic acid must be present as asparagine. The complete sequence is therefore not identical with that from bovine foetal β-chains and also bovine β-globin and sheep-β globin and contains one residue of leucine less at the *N*-terminal end in the sequence

Leu-Gly-Asn-Val-Leu-Val-Val-Val-Leu-Ala-Arg

The presence of four leucine residues in this peptide was possible and expected from the sequences of these other β-globins. The insoluble nature of the peptide is a major problem in obtaining it pure and free from partial digestion products. Nevertheless several preparations made from β-globin, CNBr2, and citraconyl-CNBr2 using two different batches of trypsin (treated with the inhibitor for chymotrypsin, tosylphenylalanylchloromethane) consistently showed a composition corresponding to that of the sequence shown.

βTp12C

The sequence of this peptide is

His-Phe-Ser-Lys

The histidine was identified as its PTH-derivative.

βTp13

This peptide gave a positive reaction with Ehrlich's reagent for tryptophan. The dansyl–Edman procedure gave the sequence

Glx-Phe-Thr-Pro-Glx-Ala-Glx-Ala-Ala-Trp-*Glx*-Lys

The tryptophan was placed by difference since no major dansyl amino acid was detected at step 10. Supporting evidence for this position came from a basic peptide, analysing as $Ala_2Trp_1Glu_1Lys_1$, which was present in a peptide map of the tryptic digest from CNBr2. Its presence was presumably due to some chymotryptic activity in the TPCK-trypsin and a longer digestion time (20 hr).

Thermolysin digestion gave two main peptides, one acidic (Thr,Glu_2,Pro, Ala,Phe) and one which was basic at ph 6·4 (Glu,Ala_2,Trp,Lys). Some smaller peptides isolated in lesser amounts assisted in determining which of the residues were glutamic acid and which were glutamine. The intact peptide had a net charge of −1. The minor peptides identified were *Ala*-Gln, neutral at pH 6·4; Ala-Trp; *Gln*-Lys, basic at pH 6·4; as well as free alanine.

The complete sequence is shown below together with peptide bonds hydrolysed by thermolysin

Glu-Phe-Thr-Pro-Glu-Ala-Gln-Ala-Ala-Trp-*Gln*-Lys

↑ ↑ ↑ ↑

βTp14

This peptide gave a clear sequence by the dansyl–Edman procedure except for steps where histidine became *N*-terminal. The position of these residues was confirmed by the composition of peptides obtained by hydrolysis with thermolysin as shown below

Leu-Val-Ser-Gly-Val-Ser-His-*Ala-Leu-Ala*-His-Lys

↑ ↑

βTp15

This is the *C*-terminal peptide since it contains no lysine or arginine. It is the same as the *C*-terminal peptide of all known β-chains with the sequence *Tyr*-His.

Complete Sequence

The complete amino acid sequence of the β-IB chain is shown in Table 4.

IV. Discussion

The length of the peptide chain in the β-globin of the echidna appears to be identical with that of other β-globin chains although a surprising change was detected in the G helix. An asparaginyl residue is present at residue 104 (G6) where a basic amino acid residue is usually present and a basic amino acid is present in the next position G7, residue 105, usually occupied by a leucine residue in all known β-chains (Dayhoff 1969). The effect of this variation on the folded structure is presumably small and an evolutionary change resulting in the presence of an asparagine residue in place of the usual basic amino acid residue has been accompanied by another mutation in the adjacent residue to a basic amino acid. Apart from this change the hydrophobic character of the G helix and the strong homology with the amino acid sequence and contact sites of other haemoglobin chains are preserved. Perutz, Kendrew, and Watson (1965) have listed the G7 residue as invariably nonpolar but in a surface crevice with no visible function. Since the substitution of an asparagine residue for an arginine residue requires a minimum of two base changes in a nucleotide sequence it is more likely that the change occurred from an ancestral

TABLE 4

COMPLETE AMINO ACID SEQUENCE OF THE β-IB CHAIN OF ECHIDNA HAEMOGLOBIN

Residues are numbered from the *N*-terminal residue and the positions of cleavage by trypsin are indicated by arrows. The 31 residues which differ from those present in human β-chain are shown in capital letters

```
                5                  10                  15                  20                  25                  30
Val-His-Leu-SER-GLY-SER-Glu-Lys-THR-Ala-Val-Thr-ASN-Leu-Trp-Gly-HIS-Val-Asn-Val-ASN-Glu-LEU-Gly-Gly-Glu-Ala-Leu-Gly-Arg-
          βTp1              ↑                                                 βTp2,3                                    ↑

               35                  40                  45                  50                  55                  60
Leu-Leu-Val-Val-Tyr-Pro-Trp-Thr-Gln-Arg-Phe-Phe-Glu-Ser-Phe-Gly-Asp-Leu-Ser-SER-ALA-Asp-Ala-Val-Met-Gly-Asn-ALA-Lys-Val-Lys-
          βTp4              ↑                                                 βTp5                                    ↑ βTp6 ↑

          65                  70                  75                  80                  85                  90
Ala-His-Gly-ALA-Lys-Val-Leu-THR-SER-Phe-GLY-Asp-ALA-Leu-LYS-ASN-Leu-Asp-Asn-Leu-Lys-Gly-Thr-Phe-Ala-LYS-Leu-Ser-Glu-
     βTp7,8 ↑               βTp9A                    ↑          βTp9B          ↑  βTp10A          ↑          βTp10B

               95                 100                 105                 110                 115                 120
Leu-His-Cys-Asp-Lys-Leu-His-Val-Asp-Pro-Glu-Asn-Phe-ASN-ARG-Leu-Gly-Asn-Val-Leu-Val-VAL-Val-Leu-Ala-ARG-His-Phe-SER-Lys-
                    ↑          βTp11,12A                ↑                    βTp12B                    ↑  βTp12C       ↑
                    ⋮
              125                 130                 135                 140                 145
Glu-Phe-Thr-Pro-GLU-ALA-Gln-Ala-Ala-TRP-Gln-Lys-LEU-Val-SER-Gly-Val-SER-HIS-Ala-Leu-Ala-His-Lys-Tyr-His
          βTp13                          ↑                    βTp14                          ↑  βTp15
```

sequence involving a lysine residue. A possible mechanism involving single base changes could then be

			Phe–Asn–Leu				
Phe–Lys–Leu	→		*or*		→	Phe–Asn–Arg	
103	105		Phe–Lys–Arg			103	105
			103	105			

Comparison of this amino acid sequence with that of horse β-globin, which has a known three-dimensional structure (Perutz 1969) shows few substitutions in the contact sites that may have any effect on the structure. Thus there are no differences from horse β-chain in the β–α_2 contact amino acids or for β-haem contacts. For the β–α_1 interactions there are variations in three positions. These contacts are mostly non-polar and have shown more variations between haemoglobins than the β–α_2 and β–haem contacts.

The most noticeable feature of the sequence of echidna β-chain in relation to the sequence of β-chains from marsupials and mammalian species is that there are less differences between the sequences of echidna β-chains and eutherian mammal β-chains than is the case for marsupial β-chains. Thus there are 31 differences between human β-globin and echidna β-globin compared with 38 differences between human-β and kangaroo β-chain. It is therefore concluded that the sequence of echidna β-chain would not support any palaeontological estimate that gave a date for marsupial divergence more recent than that for monotremes. The amino acid sequence of the α-chain is also being investigated and the calculations based on the sequence data and the statistically constant rate of change (cf. Air *et al.* 1971) will be given in a subsequent paper.

V. Acknowledgments

This work was supported in part by the Australian Research Grants Committee. The authors are indebted to Dr. M. Griffiths of the CSIRO Division of Wild Life Research, CSIRO, Canberra, and Dr. D. W. Cooper of La Trobe University, Melbourne, for samples of echidna blood and for the helpful discussions on the electrophoretic components of haemolysates.

VI. References

Air, G. M., and Thompson, E. O. P. (1969).—*Aust. J. biol. Sci.* **22**, 1437.

Air, G. M., and Thompson, E. O. P. (1971).—*Aust. J. biol. Sci.* **24**, 75.

Air, G. M., Thompson, E. O. P., Richardson, B. J., and Sharman, G. B. (1971).—*Nature, Lond.* **229**, 391.

Ambler, R. P. (1965).—*Biochem. J.* **96**, 32P.

Babin, D. R., Schroeder, W. A., Shelton, J. R., Shelton, J. B., and Robberson, B. (1966).—*Biochemistry* **5**, 1297.

Beard, J. M., and Thompson, E. O. P. (1971).—*Aust. J. biol. Sci.* **24**, 765.

Blombäch, B., Blombäch, M., Edman, P., and Hessel, B. (1966).—*Biochim. biophys. Acta* **115**, 371.

Boyer, S. H., Hathaway, P., Pascasio, F., Bordley, J., Orton, C., and Naughton, M. A. (1967).—*J. biol. Chem.* **242**, 2211.

Clegg, J. B., Naughton, M. A., and Weatherall, D. J. (1965).—*Nature, Lond.* **207**, 945.

DAYHOFF, M. O. (1969).—"Atlas of Protein Sequence and Structure." (National Biochemical Research Foundation: Silver Spring.)
DIXON, H. B. F., and PERHAM, R. N. (1968).—*Biochem. J.* **109**, 312.
DOZY, A. M., KLEIHAUER, E. F., and HUISMAN, T. H. J. (1968).—*J. Chromat.* **32**, 723.
GERALD, P. S., and INGRAM, V. M. (1961).—*J. biol. Chem.* **236**, 2155.
GRIFFITHS, M. (1968).—"Echidnas." (Pergamon Press: Oxford.)
MATSUBARA, H., and SASAKI, R. M. (1969).—*Biochem. biophys. Res. Commun.* **35**, 175.
NEUMANN, N. P., MOORE, S., and STEIN, W. H. (1962).—*Biochemistry* **1**, 68.
PERUTZ, M. F. (1969).—*Proc. R. Soc.* B**173**, 113.
PERUTZ, M. F., KENDREW, J. C., and WATSON, H. C. (1965).—*J. Molec. Biol.* **13**, 669.
SCHROEDER, W. A., SHELTON, J. R., SHELTON, J. B., ROBBERSON, B., and BABIN, D. R. (1967).—*Archs Biochem. Biophys.* **120**, 124.
THOMPSON, E. O. P. and AIR, G. M. (1971).—*Aust. J. biol. Sci.* **24**, 1199.
THOMPSON, E. O. P., HOSKEN, R., and AIR, G. M. (1969).—*Aust. J. biol. Sci.* **22**, 449.
THOMPSON, E. O. P., and O'DONNELL, I. J. (1967).—*Aust. J. biol. Sci.* **20**, 1001.
THOMPSON, E. O. P., SLEIGH, R. W., and SMITH, M. B. (1971).—*Aust. J. biol. Sci.* **24**, 525.
ZUCKERKANDL, E., and PAULING, L. (1965).—In "Evolving Genes and Proteins." (Eds. V. Bryson and H. J. Vogel.) p. 97. (Academic Press: New York.)